

Three-dimensional single-pixel compressive reflectivity imaging based on complementary modulation

Wen-Kai Yu,^{1,2} Xu-Ri Yao,^{1,2} Xue-Feng Liu,¹ Long-Zhen Li,^{1,2} and Guang-Jie Zhai^{1,*}

¹Key Laboratory of Electronics and Information Technology for Space System, Center for Space Science and Applied Research, Chinese Academy of Sciences, Beijing 100190, China

²University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author: gjzhai@nssc.ac.cn

Received 10 October 2014; revised 4 December 2014; accepted 4 December 2014;
posted 4 December 2014 (Doc. ID 224786); published 12 January 2015

A three-dimensional (3D) imager with a single-pixel detector and complementary intensity modulation of a digital micromirror device (DMD) array, which does not rely on scene raster scanning as in light detection and ranging (LIDAR) or on a two-dimensional array of sensors as used in time-of-flight (TOF) cameras, can not only capture full-color, high-quality images of real-life objects, but also recover the depth information and 3D reflectivity of the scene, reducing the required measurement dimension as well as the complexity, and cutting the cost of the detector array down to a single unit. The imager achieves spatial resolution using compressed sensing to exploit the sparsity of the signal. The disparity maps of the scene are reconstructed using sum of absolute or squared differences to reveal the depth information. This nonscanning, low-complexity 3D reflectivity imaging prototype may be of considerable value to various computer vision applications. © 2015 Optical Society of America

OCIS codes: (110.3010) Image reconstruction techniques; (110.2990) Image formation theory; (110.1650) Coherence imaging; (200.4740) Optical processing.

<http://dx.doi.org/10.1364/AO.54.000363>

1. Introduction

Acquiring three-dimensional (3D) structure and reflectivity with an active imager has many applications. Active range acquisition systems such as light detection and ranging (LIDAR) and time-of-flight (TOF) cameras obtain the range information from a single viewpoint by measuring the time difference of arrival between a transmitted pulse and the scene reflection. In the former, transverse resolution is obtained by single-pixel devices via raster scanning [1–3]. The latter typically replaces scanning with spatially resolving detectors to acquire the depth map [4,5]. However, LIDAR cameras are limited by the scanning time, and TOF cameras also

have limitation in high-resolution array fabrication. Moreover, in both cameras, single pixel or individual pixels in the sensor array are very small and the optical flux must be distributed across the entire array, so the shot noise is significant for each pixel. Thus, we can say that they achieve high depth resolution but suffer from poor spatial resolution.

The solution is to remove the need for a spatially resolving detector as well as the scanning, and computational imaging based on random patterns is a good alternative method which recovers the image of the object by correlating the known spatial distribution of a changing speckle pattern with the total reflected (or transmitted) light intensity. This technique is called computational ghost imaging (GI) [6,7], a variant of GI. Initially, ghost images were generated from two correlated light fields and two detectors: a bucket (single-pixel) detector without

spatial resolution is positioned in the signal arm to collect the total light intensity coming from an object, while a spatial-resolution detector in the reference arm is used to record the light field which had not interacted with the object. Its first demonstration [8,9] used coincidence counts of signal-idler biphoton pairs, and then it was proved that GI is also achievable with pseudothermal [10–12] or true thermal light [13], inspiring much attention and applications such as optical encryption [14,15], correspondence imaging [16,17], lensless GI with sunlight [18], GI through turbulent atmosphere [19], or adaptive GI [20]. In order to dramatically improve the quality of ghost images, Katz *et al.* [21] experimentally demonstrated that, by utilizing compressed sensing (CS) [22–24], algorithms could get a better performance with far fewer measurements than conventional GI. This is based on the single-pixel camera scheme proposed by Baraniuk *et al.* [25], where a bucket (single-pixel) detector is located on the focal plane to collect the total light intensity. We found that only the signal is magnified dramatically while the shot noise level is kept unchanged, and thus this protocol has the maximum flux and signal-to-noise ratio, as well as sensitivity. If we changed the detector with a single-photon single-pixel detector, it will make ultraweak light detection possible. In addition, with the help of CS algorithms, images can be recovered accurately from far fewer measurements than what is usually considered necessary. Therefore, we use the single-pixel camera here instead of standard cameras.

On another note, based on this single-pixel camera scheme, Howland *et al.* [26] proposed a laser radar system for 3D imaging where transverse spatial resolution is obtained through compressive sampling without scanning or array detection. Subsequently, this protocol was improved by using parametric signal modeling to recover the set of distinct depth ranges present in the scene [27]. Recently, Sun *et al.* [28] demonstrated that the 3D spatial form of an object can also be captured by comparing the shading information in the images, which are derived from several single-pixel detectors in different locations. In their scheme, they measured the differential signals of the complementary illuminated pairs, actually normalizing the bucket signals with respect to the positive or negative intensity fluctuations averaged to 0, and then correlated them with noninverted patterns to reconstruct a ghost image. Their method was essentially the same as that of correspondence imaging [16,17], but their GI bucket values were used as a series of weighting factors instead of just 0 and 1.

Recently we have developed a novel technique, which we call complementary compressive imaging, that makes full use of both complementary reflections and uses two single-pixel detectors to dramatically improve the image quality [29]. In this work, we have extended this technique to the imaging of 3D reflectivity using only a single-pixel detector, but without time-correlated module and scanning components. The spatial resolution is generated by

random patterns and their inverse ones on a digital micromirror device (DMD). The object and the light source are fixed onto the same rotating platform to ensure that the object is always illuminated from only one direction. Rotating the platform by a small angle, two images are reconstructed by a CS algorithm as if they are taken from the different viewpoints. Using a binocular stereo algorithm allows the surface gradient and hence the 3D reflectivity to be reconstructed.

2. Experimental Setup and 2D Imaging Results

Our setup (Fig. 1) consists of a light-emitting diode (LED) lamp illuminating a 3D target; a rotating platform placed with the source and the target; a rotating set composed of a red, green, and blue filter; a DMD to perform light intensity modulation with computer-generated random one-to-one complementary binary pattern pairs; a single-pixel detector to measure the total intensity of the back-reflected light; an imaging lens; a collecting lens; and a computer to generate the random complementary patterns as well as perform 3D reconstructions of the target.

The platform ensures that the LED lamp illuminates a target in only one orientation. The object is positioned at the rotational axis of the platform, which is about 26 cm away from the working plane of the DMD. The back-reflected optical flux of the object is transmitted through a rotating filter set and focused onto the DMD, which consists of 768×1024 micromirrors, each of size $13.68 \times 13.68 \mu\text{m}^2$, via the imaging lens, then though a collecting lens,

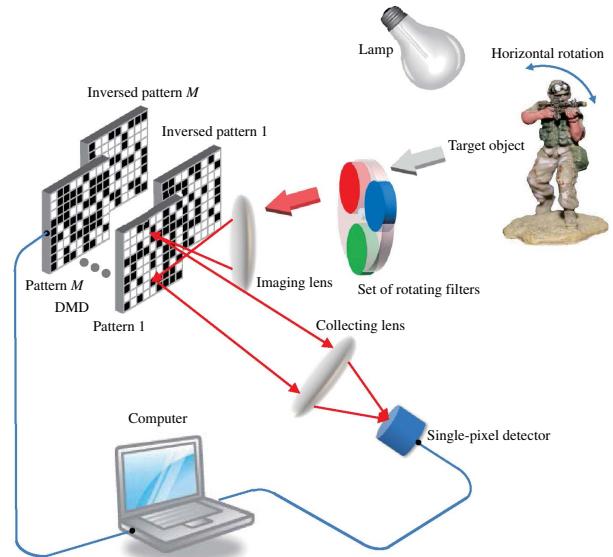


Fig. 1. Experimental setup used for 3D complementary compressive reflectivity imaging. A rotating platform placed with a LED lamp and a target turns a small angle such that two images are derived from the same single-pixel detector but appear as if they are taken from the different viewpoints, like a binocular active vision. For each rotation position, the same computer-generated random one-to-one complementary binary pattern pairs are encoded on the DMD.

the light is converged onto one spot. A photomultiplier tube (PMT) (Hamamatsu H7468-20) is used here as the bucket (single-pixel) detector for collecting and recording the total light intensity reflected from the DMD for every pattern. The integration time of PMT is set to 900 μ s, which is shorter than each flip time interval of the micromirrors, and the dead time is set to 200 μ s. The bias voltage of the PMT is 400 V. Finally, the measured signals are fed to a computer algorithm to reconstruct an image for each rotation position of the platform and then recover 3D reflectivity.

Each mirror on the DMD rotates about a hinge and can be shifted between two positions oriented at $+12^\circ$ or -12° with respect to the DMD surface, where micromirrors at $+12^\circ$ appear as bright pixels 1 and, and inversely, -12° as dark pixels 0. Note that the large operational bandwidth of the DMD (270–1800 nm) makes this technique suitable for imaging at the visible light or near-infrared light wavelengths. We set the modulation patterns of the DMD in a random binary distribution, with a black-to-white ratio close to 1:1. By changing between a binary pattern and its complementary one in turn, we modulate the light intensity at a frequency of 450 Hz. For the reason that each pattern has nearly equal numbers of black and white pixels, the difference matrix of each complementary pattern pair is a randomly distributed binary matrix taking on two values 1 and -1 , and the corresponding complementary differential bucket signal is normalized with a mean approximately 0. This has been shown to improve the image quality of the 2D reconstruction [29].

The principle behind the design of CS imaging systems can be summarized in the equation

$$y = Ax + e, \quad (1)$$

where y is a $M \times 1$ column vector of linear measurements; the measurement matrix $A \in \mathbb{R}^{M \times N}$ contains M row vectors which are reshaped from the patterns fed onto the DMD; x is an original image of interest with $p \times q$ pixels ordered in a $N \times 1$ vector, where $N = p \times q$; and e of size $M \times 1$ denotes the noise. When $M < N$, this becomes an ill-conditioned problem with infinite solutions. However, in most instances, natural signals have a sparse representation in a certain basis Ψ (e.g., Haar wavelet, discrete cosine transform, Fourier transform, or noiselet transform basis [30]). According to this prior knowledge, CS theory asserts that a small collection of nonadaptive linear measurements of a compressible signal is sufficient for perfect recovery provided the measurement matrix and representation basis are incoherent, enabling sub-Nyquist measurement. Here, we expand x on an orthogonal basis $\Psi = [\psi_1, \psi_2, \dots, \psi_N]$ as

$$x = \Psi x', \quad \text{or} \quad x = \sum_{i=1}^N x'_i \psi_i, \quad (2)$$

where $x \in \mathbb{R}^{M \times 1}$ is the coefficient sequence of x in the expansion. We say that x' is k -sparse if, at most, k of

coefficients in x' are nonzero. An empirical fact is that most images are well approximated by k -sparse expansions with k much less than the number of pixels N , and this is the reason why data compression is effective. For incoherent pairs, we only need on the order of $k \log(N/k)$ random samples. Then the problem becomes

$$y = A\Psi x' + e. \quad (3)$$

Let y' and A' denote the bucket signal and the measurement matrix, respectively. Since each complementary frame pair here appears alternately, the complementary differential bucket signal is defined as $y = y'_o - y'_{o+1}$, $o = 1, 3, 5, \dots$. Accordingly, complementary differential frames can be written as $A = A'_o - A'_{o+1}$, where inverse frames $A'_{o+1} = 1_{M \times N} - A'_o$, $1_{M \times N}$ stand for an array of all 1. Then Eq. (3) can be rewritten as

$$y'_o - y'_{o+1} = (A'_o - A'_{o+1})\Psi x' + (e_1 - e_2). \quad (4)$$

Further, the accurate recovery is achieved by solving a tractable convex optimization program [24]. Recent research [31] has proved that the use of total variation (TV) regularization instead of the l_1 term in CS problems gives a sharper recovered image by preserving the edges or boundaries more accurately, and the gradient of an image is generally sparse as well. For reconstructing an image, a solver named TVAL3 [31] is applied to this TV-based minimization model:

$$\min_x \sum_i \|D_i x\|_1 + \frac{\mu}{2} \|y - Ax\|_2^2, \quad (5)$$

where $D_i x$ is the discrete gradient vector of x at position i , D is the gradient operator, $\sum_i \|D_i x\|_1$ is the discrete TV of x , μ is a constant scalar used to balance these two terms, and $\|\cdot\|_1$ stands for l_1 norm, defined as $\|x\|_1 = \sum_{i=1}^N |x_i|$. The first term is small when $D_i x$ is sparse. The second term is small when the optimal x is consistent with Eq. (3) within a small error.

By changing the color filters, we obtain red, green, and blue component images for each rotation position of the platform via CS, as shown in Figs. 2(a)–2(c) and 2(e)–2(g). Then the colored images [see Figs. 2(d) and 2(h)] are recovered by synthesizing the three reconstructed components using multiple grayscale encoding. In order to acquire a 2D image of size 128×128 pixels, it would take $128^2 = 16384$ measurements as in LIDAR or 16,384 sensors in a TOF camera. However, in this framework, only 9856 measurements (or patterns) in total, about 60% of the total number of pixels, are used for high-quality 2D image reconstruction. In comparison, LIDAR and TOF cameras all suffer from poor spatial resolution and do not use the sparsity inherent in 2D imaging to achieve savings in number of sensors or scanning pixels. The imaging time of

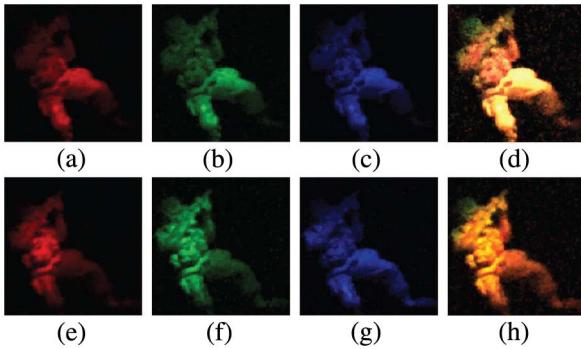


Fig. 2. Color reconstructions of a 3D target for each rotation position of the platform. (a)–(c) and (e)–(g) correspond to red, green, and blue color channels of the 2D image in the first and second positions. (d) and (h) are full-color reconstructions obtained by combining the three separate color components (a)–(c) and (e)–(g), respectively. The patterns utilized group clusters of 6×6 mirrors, so the size of one “pixel” (minimum resolution) is $82.08 \times 82.08 \mu\text{m}^2$. Since the imaging region covers 768×768 mirrors in the center of the DMD, we will finally obtain a 2D image of size 128×128 “pixels” with 255 grayscales.

our system was $9856/450 = 21.9$ s. In the future, the modulation frequency of the DMD can reach 32.5 kHz, and the detector can be replaced with an APD or counter-type PMT of GHz, so the imaging time of our system can be greatly improved. In addition, applying the single-pixel camera scheme has the great advantages of high signal-to-noise ratio and high sensitivity in large flux subsampling with only a single pixel, compared with standard cameras, although it may sacrifice some imaging time with the use of the DMD. We also find that both LIDAR and TOF cameras need a much longer integration time than our imaging system, for the reason that the optical flux must be distributed across the entire array on the imaging plane. If the total number of image pixels is large enough, the imaging time of our system will be comparable with that of the standard cameras.

3. 3D Reconstructions Using Binocular Stereo Vision

Because the object is fixed in the center of the platform, and the LED lamp is suspended in midair but with a support pole fixed onto the edge of the platform, the apparent lighting of the object depends on the illumination orientation of the light source. However, the images [Figs. 2(d) and 2(h)] are derived from the different rotation positions of the platform, and the intensity distribution of these two images is different. Normally, depth information of a scene is lost in a 2D image, but there is a stereo vision technique, named binocular stereo vision, which can extract depth information from two images taken from different viewpoints just similar to our eyes. Stereo vision algorithms can be roughly divided into feature-based and area-based algorithms. Feature-based algorithms are based on epipolar geometry, where each object point in one of the stereo images can be found on a specific line, called the epipolar line, in the other image. They use characteristics

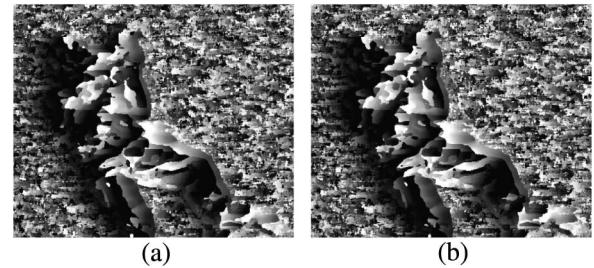


Fig. 3. 3D reconstructions using binocular stereo vision: disparity maps retrieved by (a) SAD-based and (b) SSD-based stereo-matching algorithms.

in the images such as edges or corners, comparing their similarities to reveal the depth of the scene and then to build the disparity map computing the displacement between those features. Area-based algorithms match blocks of pixels to find correspondences in the images. Common methods for the matching in area-based algorithms aggregate the sum of absolute or squared differences (SAD/SSD), over a window. These methods can be implemented efficiently using filters, e.g., Laplacian of Gaussian (LoG) and mean filters, for removing noise and changes in bias. There also exist various other algorithms for area-based matching, but most of them are computationally too expensive. Given that SAD and SSD are easy to implement and use less computing power, we use them here for 3D reconstructions. SAD and SSD are defined as

$$\text{SAD} = \sum_{i=-L}^L \sum_{j=-L}^L |I_{r+i,c+j} - I'_{r'+i,c'+j}|, \quad (6)$$

$$\text{SSD} = \sum_{i=-L}^L \sum_{j=-L}^L |I_{r+i,c+j} - I'_{r'+i,c'+j}|^2, \quad (7)$$

where $L = (s - 1)/2$, while I is the primary and I' is the secondary image being matched against each other, having r, c or r', c' respectively as the center coordinates of the current block. We perform the matching using the image of the first rotation position as the primary one. The block size s affects the quality of the disparity map.

After enlarging 2D images shown in Fig. 2 fourfold on their scales to get more computable pixels, we calculated the SAD and SSD with a window size of 9×9 pixels, and finally captured the depth images as illustrated in Fig. 3. As defined in Eqs. (6) and (7), the SSD algorithm performs only a little better than the SAD algorithm, with a difference of the square operation. Thus, the performance of these two algorithms as shown in Fig. 3 is similar.

4. Conclusion

In conclusion, we have experimentally demonstrated a 3D compressive reflectivity imaging system with only a single-pixel detector and complementary

intensity modulation performed by a DMD. The system uses a convex optimization algorithm to reconstruct full-color 2D images with the sparsest coefficients represented in some basis followed by calculating SAD/SSD to reveal the spatial structure of the depth maps. Unlike LIDAR and TOF cameras, it uses the sparsity inherent in a 2D image to achieve savings in number of scanning measurements as in LIDAR or sensors in TOF, but with high-quality spatial resolution. The use of a rotating platform positioned with a LED lamp in the edge and a 3D target in the center ensures that the LED lamp illuminates a target in a fixed direction. By rotating the platform, we will acquire two full-color images from different viewpoints with different shading distribution. Utilizing binocular stereo vision makes it feasible to extract depth information. An important difference between our technique and LIDAR/TOF cameras is that only a single detector is used to resolve spatial resolution and recover the disparity maps of the object, removing the need of scanning or array detector and reducing the required measurement dimension, complexity, and cost, something that can be crucial for various computer vision applications.

We warmly acknowledge Ling-An Wu for providing helpful suggestions. This work was supported by the National Key Scientific Instrument and Equipment Development Project of China (Grant No. 2013YQ030595), and the National High Technology Research and Development Program of China (Grant No. 2013AA122902).

References

1. F. Blais, "Review of 20 years of range sensor development," *J. Electron. Imaging* **13**, 231–240 (2004).
2. R. Lamb and G. Buller, "Single-pixel imaging using 3d scanning time-of-flight photon counting," SPIE Newsroom, 2010, <http://spie.org/x39222.xml>.
3. B. Schwarz, "LIDAR: mapping the world in 3D," *Nat. Photonics* **4**, 429–430 (2010).
4. S. Foix, G. Alenyà, and C. Torras, "Lock-in time-of-flight (ToF) cameras: a survey," *IEEE Sens. J.* **11**, 1917–1926 (2011).
5. A. Medina, F. Gayá, and F. del Pozo, "Compact laser radar and three-dimensional camera," *J. Opt. Soc. Am. A* **23**, 800–805 (2006).
6. J. H. Shapiro, "Computational ghost imaging," *Phys. Rev. A* **78**, 061802 (2008).
7. Y. Bromberg, O. Katz, and Y. Silberberg, "Ghost imaging with a single detector," *Phys. Rev. A* **79**, 053840 (2009).
8. T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Phys. Rev. A* **52**, R3429(R) (1995).
9. D. V. Strekalov, A. V. Sergienko, D. N. Klyshko, and Y. H. Shih, "Observation of two-photon "ghost" interference and diffraction," *Phys. Rev. Lett.* **74**, 3600–3603 (1995).
10. A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato, "Ghost imaging with thermal light: comparing entanglement and classical correlation," *Phys. Rev. Lett.* **93**, 093602 (2004).
11. F. Ferri, D. Magatti, A. Gatti, M. Bache, E. Brambilla, and L. A. Lugiato, "High-resolution ghost image and ghost diffraction experiments with thermal light," *Phys. Rev. Lett.* **94**, 183602 (2005).
12. A. Gatti, M. Bache, D. Magatti, E. Brambilla, F. Ferri, and L. A. Lugiato, "Coherent imaging with pseudo-thermal incoherent light," *J. Mod. Opt.* **53**, 739–760 (2006).
13. D. Zhang, Y. H. Zhai, L. A. Wu, and X. H. Chen, "Correlated two-photon imaging with true thermal light," *Opt. Lett.* **30**, 2354–2356 (2005).
14. S. Li, X. R. Yao, W. K. Yu, L. A. Wu, and G. J. Zhai, "High-speed secure key distribution over an optical network based on computational correlation imaging," *Opt. Lett.* **38**, 2144–2146 (2013).
15. W. K. Yu, S. Li, X. R. Yao, X. F. Liu, L. A. Wu, and G. J. Zhai, "Protocol based on compressed sensing for high-speed authentication and cryptographic key distribution over a multiparty optical network," *Appl. Opt.* **52**, 7882–7888 (2013).
16. L. A. Wu and K. H. Luo, "Two-photon imaging with entangled and thermal light," *AIP Conf. Proc.* **1384**, 223–228 (2011).
17. K. H. Luo, B. Q. Huang, W. M. Zheng, and L. A. Wu, "Nonlocal imaging by conditional averaging of random reference measurements," *Chin. Phys. Lett.* **29**, 074216 (2012).
18. X. F. Liu, X. H. Chen, X. R. Yao, W. K. Yu, G. J. Zhai, and L. A. Wu, "Lensless ghost imaging with sunlight," *Opt. Lett.* **39**, 2314–2317 (2014).
19. J. Cheng, "Ghost imaging through turbulent atmosphere," *Opt. Express* **17**, 7916–7921 (2009).
20. W. K. Yu, M. F. Li, X. R. Yao, X. F. Liu, L. A. Wu, and G. J. Zhai, "Adaptive compressive ghost imaging based on wavelet trees and sparse representation," *Opt. Express* **22**, 7133–7144 (2014).
21. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.* **95**, 131110 (2009).
22. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
23. D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
24. E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians* (European Mathematical Society, 2006), Vol. **3**, pp. 1433–1452.
25. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.* **25**(2), 83–91 (2008).
26. G. A. Howland, P. B. Dixon, and J. C. Howell, "Photon-counting compressive sensing laser radar for 3D imaging," *Appl. Opt.* **50**, 5917–5920 (2011).
27. A. Colaço, A. Kirmani, G. A. Howland, J. C. Howell, and V. K. Goyal, "Compressive depth map acquisition using a single photon-counting detector: parametric signal processing meets sparsity," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012), pp. 96–102.
28. B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. J. Padgett, "3D computational imaging with single-pixel detectors," *Science* **340**, 844–847 (2013).
29. W. K. Yu, X. F. Liu, X. R. Yao, C. Wang, Y. Zhai, and G. J. Zhai, "Complementary compressive imaging for the telescopic system," *Sci. Rep.* **4**, 5834 (2014).
30. S. Olivas, Y. Rachlin, L. Gu, B. Gardiner, R. Dawson, J. Laine, and J. Ford, "Characterization of a compressive imaging system using laboratory and natural light scenes," *Appl. Opt.* **52**, 4515–4526 (2013).
31. C. B. Li, "An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing," Masters of Science thesis (Rice University, 2010).